

Injectivity and weak*-to-weak continuity suffice for convergence rates in ℓ^1 -regularization

JENS FLEMMING and DANIEL GERTH *

January 16, 2017

Key words: linear ill-posed problem, sparsity promoting regularization, Tikhonov regularization, source condition, variational source condition, convergence rates

MSC: 65J20, 47A52

Abstract

We show that the convergence rate of ℓ^1 -regularization for linear ill-posed equations is always $\mathcal{O}(\delta)$ if the exact solution is sparse and if the considered operator is injective and weak*-to-weak continuous. Under the same assumptions convergence rates in case of non-sparse solutions are proven. The results base on the fact that certain source-type conditions used in the literature for proving convergence rates are automatically satisfied.

1 Setting and main theorem

Let $A : \ell^1 \rightarrow Y$ be a bounded linear operator mapping absolutely summable real sequences into a real Banach space Y . For solving the possibly ill-posed equation

$$Ax = y^\dagger, \quad x \in \ell^1, \quad (1.1)$$

we consider ℓ^1 -regularization. That is, given noisy data y^δ in Y with

$$\|y^\delta - y^\dagger\|_Y \leq \delta$$

*Chemnitz University of Technology, Faculty of Mathematics, D-09107 Chemnitz, Germany, jens.flemming@mathematik.tu-chemnitz.de, daniel.gerth@mathematik.tu-chemnitz.de.

for some positive δ , we solve

$$T_\alpha^\delta(x) := \|Ax - y^\delta\|_Y^p + \alpha\|x\|_{\ell^1} \rightarrow \min_{x \in \ell^1}. \quad (1.2)$$

Here $\alpha > 0$ is the regularization parameter controlling the influence of the penalty term and $p > 1$ is some exponent which can be used to simplify numerical minimization.

By x_α^δ we denote a minimizer of T_α^δ . Throughout this article we assume that (1.1) has a solution x^\dagger in ℓ^1 and the aim is to find asymptotic estimates (convergence rates) for the solution error $\|x_\alpha^\delta - x^\dagger\|_{\ell^1}$ in terms of the noise level δ . To ensure convergence of x_α^δ to x^\dagger we have to choose α in the right way depending on δ and y^δ . In the following we restrict our attention to *a priori* choices $\alpha = \alpha(\delta)$ and to the *discrepancy principle*

$$\delta \leq \|Ax_\alpha^\delta - y^\delta\|_Y \leq \tau\delta \quad (1.3)$$

with $\tau \geq 1$. The later means that $\alpha = \alpha(\delta, y^\delta)$ is chosen such that the corresponding discrepancy is close to δ . Both parameter choice methods are well-known and we refer to [5, 16] for details.

Since ℓ^1 is the dual space of c_0 , the space of sequences converging to zero, we have the notion of weak* convergence in ℓ^1 at hand. If A is sequentially weak*-to-weak continuous, then T_α^δ has minimizers and ℓ^1 -regularization (1.2) is a stable and convergent method. In addition, the minimizers are sparse, that is they have only finitely many non-zero components. These results can be found, e. g., in [7, Section 2].

Convergence rates for ℓ^1 -regularization in infinite dimensions were obtained at first in [4, Proposition 4.7] based on smoothing properties of A . In [2] rates were obtained if the canonical basis of ℓ^1 belongs to the range of the adjoint A^* and in [11] rates were shown for more general penalties in the Tikhonov functional based only on an injectivity-type assumption, but ℓ^q -regularization is only covered if $q < 1$. Further rates results can be found in [12, 14] based on a Banach space source condition. The mentioned convergence rates results only hold if the exact solution x^\dagger is sparse. First rates results for non-sparse solutions were presented in [3] with the same range condition for the canonical basis as in [2]. Under weaker assumptions same results were proven in [8, 9].

Now we state our main result which shows that next to injectivity and sequential weak*-to-weak continuity of A no further assumptions like source conditions are needed to prove convergence rates for sparse (see corollary below) and non-sparse solutions. In fact, sequential weak*-to-weak continuity does not restrict the scope of application because to our best knowledge

this property is the weakest assumption ensuring existence of regularized solutions as well as stability and convergence of ℓ^1 -regularization (cf. [7, 16]). Thus, the only essential assumption is injectivity of A .

Theorem 1. *Let $A : \ell^1 \rightarrow Y$ be an injective and sequentially weak*-to-weak continuous bounded linear operator and denote by $x^\dagger \in \ell^1$ the solution of (1.1). Then there are a continuous, concave and monotonically increasing function $\varphi : [0, \infty) \rightarrow [0, \infty)$ with $\varphi(0) = 0$ and a constant c such that*

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^1} \leq c\varphi(\delta) \quad \text{for all } \delta > 0,$$

if the regularization parameter α is chosen a priori $\alpha \sim \frac{\delta^p}{\varphi(\delta)}$ or by the discrepancy principle (1.3). Further, there is always a monotonically increasing sequence $(\gamma_n)_{n \in \mathbb{N}}$ of positive numbers such that φ can be chosen

$$\varphi(t) := 2 \inf_{n \in \mathbb{N}} \left(\sum_{k=n+1}^{\infty} |x_k^\dagger| + \gamma_n t \right). \quad (1.4)$$

The constant c is independent of x^\dagger and φ .

The proof will be given in Sections 3 and 4, where also the constant c is made explicit. In case of sparse solutions the theorem specializes to the following result.

Corollary 2. *Let $A : \ell^1 \rightarrow Y$ be an injective and sequentially weak*-to-weak continuous bounded linear operator and denote by $x^\dagger \in \ell^1$ the solution of (1.1). If x^\dagger is sparse, then there is a constant c such that*

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^1} \leq c\delta \quad \text{for all } \delta > 0,$$

if the regularization parameter α is chosen a priori $\alpha \sim \delta^{p-1}$ or by the discrepancy principle (1.3). The constant c depends on the number of non-vanishing components of x^\dagger .

Proof. Let $x_k^\dagger = 0$ for $k > m$. Then φ in Theorem 1 satisfies

$$\varphi(t) \leq 2 \sum_{k=m+1}^{\infty} |x_k^\dagger| + 2\gamma_m t = 2\gamma_m t$$

and the corresponding error estimate reduces to

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^1} \leq 2c\gamma_m \delta.$$

□

In the next section we discuss the relation of the range of A^* to smoothness properties of basis elements. Then we go on to the proof of our main theorem in Sections 3 and 4.

2 The range of A^* and basis smoothness

Denote by $\ell^\infty = (\ell^1)^*$ the space of bounded sequences and by $A^* : Y^* \rightarrow \ell^\infty$ the adjoint of A . The range $\mathcal{R}(A^*)$ of A^* played a crucial role in several results on convergence rates for ℓ^1 -regularization and in the present section we discuss some known results on its structure from the literature.

Denoting by $(e^{(k)})_{k \in \mathbb{N}}$ the canonical basis of ℓ^1 , the following lemma has been proven in [7].

Lemma 3. *The following assertions are equivalent:*

- (i) *A is sequentially weak*-to-weak continuous.*
- (ii) $\mathcal{R}(A^*) \subseteq c_0$.
- (iii) $(Ae^{(k)})_{k \in \mathbb{N}}$ converges weakly to zero.

Proof. See [7, Lemma 2.1]. □

On the one hand the lemma shows that sequential weak*-to-weak continuity can be reformulated as a property of the range $\mathcal{R}(A^*)$. On the other hand, item (iii) in the lemma is obviously satisfied if A has a bounded extension to some ℓ^q -space with $q > 1$. Thus, restriction to sequentially weak*-to-weak continuous operators is a very weak restriction. But note that for $Y = \ell^1$ the identity mapping is a simple example of a not sequentially weak*-to-weak continuous operator.

First results on convergence rates for ℓ^1 -regularization if solutions are not sparse in [3] were based on the assumption that

$$e^{(k)} \in \mathcal{R}(A^*) \quad \text{for all } k \in \mathbb{N}. \quad (2.1)$$

To our best knowledge such a condition appeared first for nonlinear operators in [10]. In the present paper we show that the slightly weaker condition

$$e^{(k)} \in \overline{\mathcal{R}(A^*)} \quad \text{for all } k \in \mathbb{N} \quad (2.2)$$

automatically holds for all injective, sequentially weak*-to-weak continuous operators and suffices, in combination with sequential weak*-to-weak continuity, to obtain convergence rates. Here and in the whole paper an overlined subset of ℓ^∞ denotes the closure of this set with respect to the ℓ^∞ -norm.

In [3, Proposition 2.4] it has been observed that $\overline{\mathcal{R}(A^*)} = c_0$ if $\|Ae^{(k)}\|_Y$ converges to zero. The same equality will be obtained in the present paper, but under the weaker assumption that A is weak*-to-weak continuous.

Typically one has a decomposition

$$A = \tilde{A} \circ L$$

where $\tilde{A} : \tilde{X} \rightarrow Y$ maps from some Banach space \tilde{X} into Y and $L : \ell^1 \rightarrow \tilde{X}$ is a synthesis operator with respect to some Schauder basis $(v_k)_{k \in \mathbb{N}}$, that is,

$$Lx := \sum_{k=1}^{\infty} x_k v_k.$$

In practice \tilde{X} is often a Hilbert space and $(v_k)_{k \in \mathbb{N}}$ is an orthonormal basis. Then it is easy to see that (2.1) holds if and only if $v_k \in \mathcal{R}(\tilde{A}^*)$ for all k . If $\tilde{X} = \ell^p$ a similar result can be obtained and for a general Banach space \tilde{X} one has to switch to biorthogonal systems in \tilde{X} and \tilde{X}^* .

It had not been clear whether condition (2.1) holds for larger classes of operators. An affirmative was given in [1] in the case that \tilde{X} is a Hilbert space with a separable and dense linear subspace V , such that (V, \tilde{X}, V^*) forms a Gelfand triple. Among other examples it was shown that the Radon transform possesses this property, showing that a large class of in particular practical problems fulfill (2.1).

On the other hand, a negative example was constructed in [8] showing that already a certain bidiagonal operator does not fulfill (2.1). In order to overcome this deficiency, a weaker assumption for obtaining convergence rates was introduced which in principle states that there are elements $\eta \in Y^*$ such that each basis element $e^{(k)}$ can be approximated via $A^* \eta$ with $[A^* \eta]_l = e_l^{(k)}$ for $l \leq k$ and $[A^* \eta]_l$ sufficiently small for all $l > k$. For proving our main theorem we shall employ a very similar condition which in turn is a variant of an assumption that was used in [9], implying the one in [8]. In order to formulate it we introduce the projectors

$$P_n : \ell^\infty \rightarrow \ell^\infty, \quad P_n x := (x_1, \dots, x_n, 0, \dots) \quad (2.3)$$

as the cut-off after the n -th entry.

Condition 4. There exist a real sequence $(\gamma_n)_{n \in \mathbb{N}}$ and a constant $\mu \in [0, 1)$ such that for each $n \in \mathbb{N}$ and each $\xi \in \ell^\infty$ with

$$\xi_k \begin{cases} \in \{-1, 0, 1\}, & \text{if } k \leq n, \\ = 0, & \text{if } k > n \end{cases}$$

there exists some $\eta = \eta(\mu, n, \xi)$ in Y^* such that

- (i) $P_n A^* \eta = \xi$,
- (ii) $|[(I - P_n) A^* \eta]_k| \leq \mu$ for all $k > n$,
- (iii) $\|\eta\|_{Y^*} \leq \gamma_n$.

The existence of $\eta = \eta(\xi)$ in the condition depends heavily on the constants μ and γ_n . The interplay of these two constants seems to be rather complicated, but the proof of the main theorem will show that fixing some $\mu \in (0, 1)$ one always finds a sequence $(\gamma_n)_{n \in \mathbb{N}}$ such that the condition holds. The growth of the γ_n has influence on the convergence rate of ℓ^1 -regularization and it is not clear how to choose μ to make this growth as slow as possible.

3 Proof part I: variational source condition

To prove Theorem 1 we start with the properties of φ defined by (1.4). As an infimum of affine functions it is concave and upper semi-continuous. Concavity implies continuity on $(0, \infty)$ and from $\varphi(0) = 0$, non-negativity and upper semi-continuity we obtain continuity of φ on $[0, \infty)$. Monotonicity of φ follows from monotonicity of $(\gamma_n)_{n \in \mathbb{N}}$.

The major part of the proof of Theorem 1 is to show that Condition 4 holds. Then we can refer to [9, Theorem 2.2] to obtain a variational source condition (or variational inequality)

$$\beta \|x - x^\dagger\|_{\ell^1} \leq \|x\|_{\ell^1} - \|x^\dagger\|_{\ell^1} + \varphi(\|Ax - Ax^\dagger\|_Y) \quad \text{for all } x \in \ell^1. \quad (3.1)$$

with some constant $\beta \in (0, 1]$. This variational source condition is known to imply the asserted convergence rate, see [6, 13].

We now prove validity of Condition 4 and derive a variational source condition. In the next section, for the reader's convenience, we present the remaining steps to obtain the rates result in our notation.

Definition 5. Let X be some Banach space and let U and V be subspaces of X and X^* , respectively. The annihilator of $U \subset X$ in X^* is

$$U^\perp := \{\xi \in X^* : \langle \xi, x \rangle_{X^* \times X} = 0 \text{ for all } x \in U\}$$

and the annihilator of $V \subset X^*$ in X is

$$V_\perp := \{x \in X : \langle \xi, x \rangle_{X^* \times X} = 0 \text{ for all } \xi \in V\}.$$

With the help of annihilators we can carry over well known relations between null spaces and ranges of A and A^* in Hilbert spaces to Banach spaces. We need the following relation.

Lemma 6. *Denoting by $A^{**} := (A^*)^* : (\ell^\infty)^* \rightarrow Y^{**}$ we have*

$$\overline{\mathcal{R}(A^*)} = \mathcal{N}(A^{**})_\perp.$$

Proof. See, e. g., [15, Lemma 3.1.16 and Proposition 1.10.15(c)]. □

To exploit the lemma we need information about the structure of $(\ell^\infty)^*$ which is not the same as ℓ^1 but a strictly larger space, because ℓ^1 is not reflexive. We have the following very useful characterization of $(\ell^\infty)^*$, which is a special case of [17, Theorem 2.14].

Lemma 7. *Each element of $(\ell^\infty)^*$ is the sum of an element of ℓ^1 and an element of c_0^\perp , that is,*

$$(\ell^\infty)^* = \ell^1 \oplus c_0^\perp.$$

Proof. Let $u \in (\ell^\infty)^*$. Set

$$x_k := \langle u, e^{(k)} \rangle_{(\ell^\infty)^* \times \ell^\infty}.$$

Then $x = (x_k)_{k \in \mathbb{N}} \in \ell^1$ because

$$\begin{aligned} \sum_{k=1}^n |x_k| &= \sum_{k=1}^n (\operatorname{sgn} x_k) \langle u, e^{(k)} \rangle_{(\ell^\infty)^* \times \ell^\infty} = \left\langle u, \sum_{k=1}^n (\operatorname{sgn} x_k) e^{(k)} \right\rangle_{(\ell^\infty)^* \times \ell^\infty} \\ &\leq \|u\|_{(\ell^\infty)^*}. \end{aligned}$$

It remains to show $u - x \in c_0^\perp$. Indeed, for each $\xi \in c_0$ we have

$$\begin{aligned} \langle u - x, \xi \rangle_{(\ell^\infty)^* \times \ell^\infty} &= \lim_{n \rightarrow \infty} \left\langle u - x, \sum_{k=1}^n \xi_k e^{(k)} \right\rangle_{(\ell^\infty)^* \times \ell^\infty} \\ &= \lim_{n \rightarrow \infty} \sum_{k=1}^n \left(\xi_k \langle u, e^{(k)} \rangle_{(\ell^\infty)^* \times \ell^\infty} - \xi_k \langle x, e^{(k)} \rangle_{\ell^1 \times \ell^\infty} \right) \\ &= 0. \end{aligned}$$

□

Combining this result with Lemma 6 yields a full characterization of $\overline{\mathcal{R}(A^*)}$.

Proposition 8. *Let A be injective and sequentially weak*-to-weak continuous. Then*

$$\overline{\mathcal{R}(A^*)} = c_0.$$

Proof. From Lemma 7 we know that A^{**} maps $\ell^1 \oplus c_0^\perp$ into Y^{**} . On the one hand, for each $x \in \ell^1$ and each $\eta \in Y^*$ we see

$$\begin{aligned} \langle A^{**}x, \eta \rangle_{Y^{**} \times Y^*} &= \langle x, A^*\eta \rangle_{(\ell^\infty)^* \times \ell^\infty} = \langle x, A^*\eta \rangle_{\ell^1 \times \ell^\infty} = \langle Ax, \eta \rangle_{Y \times Y^*} \\ &= \langle Ax, \eta \rangle_{Y^{**} \times Y^*}, \end{aligned}$$

that is, $A^{**}|_{\ell^1} = A$. If A is injective we have $\mathcal{N}(A|_{\ell^1}) = \{0\}$. On the other hand, for each $u \in c_0^\perp$ and each $\eta \in Y^*$ we see

$$\langle A^{**}u, \eta \rangle_{Y^{**} \times Y^*} = \langle u, A^*\eta \rangle_{(\ell^\infty)^* \times \ell^\infty} = 0$$

because $A^*\eta \in \mathcal{R}(A^*) \subseteq c_0$ as a consequence of weak*-to-weak continuity (cf. [7, Lemma 2.1]). Thus, $A^{**}|_{c_0^\perp} = 0$ and together with $\mathcal{N}(A|_{\ell^1}) = \{0\}$ it is $\mathcal{N}(A^{**}) = c_0^\perp$.

With Lemma 6 we now obtain

$$\overline{\mathcal{R}(A^*)} = \mathcal{N}(A^{**})_\perp = (c_0^\perp)_\perp = c_0,$$

where the last equality is a consequence of the Hahn-Banach Theorem (cf. [15, Proposition 1.10.15]). \square

The converse, that $\overline{\mathcal{R}(A^*)} = c_0$ implies injectivity, is in general not true, because one can show that A is injective if and only if the weak*-closure of $\mathcal{R}(A^*)$ coincides with ℓ^∞ (see [15, Theorem 3.1.17(a)]).

Verification of Condition 4 will be completed by a corollary of the following proposition.

Proposition 9. *Let $\varepsilon > 0$ and let $n \in \mathbb{N}$. Then for each $\xi \in c_0$ there exists $\tilde{\xi} \in \mathcal{R}(A^*)$ such that*

$$\tilde{\xi}_k = \xi_k \quad \text{for } k \leq n \quad \text{and} \quad |\tilde{\xi}_k - \xi_k| \leq \varepsilon \quad \text{for } k > n.$$

Proof. We proof the proposition by induction with respect to n . For $\xi \in c_0$ set

$$\xi^+ := (\xi_1 + \varepsilon, \xi_2, \xi_3, \dots) \quad \text{and} \quad \xi^- := (\xi_1 - \varepsilon, \xi_2, \xi_3, \dots).$$

By Proposition 8 we find $\tilde{\xi}^+ \in \mathcal{R}(A^*)$ and $\tilde{\xi}^- \in \mathcal{R}(A^*)$ with

$$\|\tilde{\xi}^+ - \xi^+\|_{\ell^\infty} \leq \varepsilon \quad \text{and} \quad \|\tilde{\xi}^- - \xi^-\|_{\ell^\infty} \leq \varepsilon.$$

Consequently, $\tilde{\xi}_1^+ \geq \xi_1 \geq \tilde{\xi}_1^-$ and $|\tilde{\xi}_k^+ - \xi_k| \leq \varepsilon$ as well as $|\tilde{\xi}_k^- - \xi_k| \leq \varepsilon$ for $k > 1$. Thus we find a convex combination $\tilde{\xi}$ of $\tilde{\xi}^+$ and $\tilde{\xi}^-$ such that $\tilde{\xi}_1 = \xi_1$. This $\tilde{\xi}$ obviously also satisfies $|\tilde{\xi}_k - \xi_k| \leq \varepsilon$ for $k > 1$, which proves the proposition for $n = 1$.

Now let the proposition be true for $n = m$. We prove it for $n = m + 1$. Let $\xi \in c_0$ and set

$$\begin{aligned}\xi^+ &:= (\xi_1, \dots, \xi_m, \xi_{m+1} + \varepsilon, \xi_{m+2}, \xi_{m+3}, \dots), \\ \xi^- &:= (\xi_1, \dots, \xi_m, \xi_{m+1} - \varepsilon, \xi_{m+2}, \xi_{m+3}, \dots).\end{aligned}$$

By the induction hypothesis we find $\tilde{\xi}^+ \in \mathcal{R}(A^*)$ and $\tilde{\xi}^- \in \mathcal{R}(A^*)$ with

$$\tilde{\xi}_k^+ = \xi_k = \tilde{\xi}_k^- \quad \text{for } k \leq m$$

and

$$|\tilde{\xi}_k^+ - \xi_k^+| \leq \varepsilon \quad \text{and} \quad |\tilde{\xi}_k^- - \xi_k^-| \leq \varepsilon \quad \text{for } k > m.$$

Consequently, $\tilde{\xi}_{m+1}^+ \geq \xi_{m+1} \geq \tilde{\xi}_{m+1}^-$ and $|\tilde{\xi}_k^+ - \xi_k| \leq \varepsilon$ as well as $|\tilde{\xi}_k^- - \xi_k| \leq \varepsilon$ for $k > m + 1$. Thus we find a convex combination $\tilde{\xi}$ of $\tilde{\xi}^+$ and $\tilde{\xi}^-$ such that $\tilde{\xi}_{m+1} = \xi_{m+1}$. This $\tilde{\xi}$ obviously also satisfies $\tilde{\xi}_k = \xi_k$ for $k < m + 1$ and $|\tilde{\xi}_k - \xi_k| \leq \varepsilon$ for $k > m + 1$, which proves the proposition for $n = m + 1$. \square

Corollary 10. *Let A be injective and sequentially weak*-to-weak continuous. Then for each $\mu \in (0, 1)$ there is a sequence $(\gamma_n)_{n \in \mathbb{N}}$ such that Condition 4 is satisfied.*

Proof. Fix μ and n and take some ξ as described in Condition 4. By Proposition 9 with $\varepsilon := \mu$ there exists some η such that $A^*\eta (= \tilde{\xi}$ in the proposition) satisfies items (i) and (ii) in the condition.

The set of all ξ to be considered in Condition 4 for fixed n is a bounded subset of a finite-dimensional subspace of ℓ^∞ and the set of corresponding $\eta = \eta(\xi)$ is contained in the preimage of this finite-dimensional subset with respect to the linear mapping A^* . Thus, the set of η is bounded, too. Choosing γ_n to be this bound we automatically satisfy item (iii) in the condition. \square

To obtain a variational source condition (3.1) from Condition 4, which is always satisfied under our standing assumptions, we use the estimates from [9, proof of Theorem 2.2].

Corollary 11. *Let A be injective and sequentially weak*-to-weak continuous and let μ and $(\gamma_n)_{n \in \mathbb{N}}$ be as in Condition 4. Then a variational source condition (3.1) with $\beta = \frac{1-\mu}{1+\mu}$ and φ given by (1.4) is fulfilled.*

Proof. Fix $n \in \mathbb{N}$ and $x \in \ell^1$ and let $\xi := \text{sgn } P_n(x - x^\dagger) \in \ell^\infty$ be the sequence of signs of $P_n(x - x^\dagger)$. Then by Condition 4 there is some η such that

$$\begin{aligned} \|P_n(x - x^\dagger)\|_{\ell^1} &= \langle \xi, x - x^\dagger \rangle_{\ell^\infty \times \ell^1} = \langle P_n A^* \eta, x - x^\dagger \rangle_{\ell^\infty \times \ell^1} \\ &= \langle P_n A^* \eta - A^* \eta, x - x^\dagger \rangle_{\ell^\infty \times \ell^1} + \langle A^* \eta, x - x^\dagger \rangle_{\ell^\infty \times \ell^1} \\ &= -\langle (I - P_n) A^* \eta, (I - P_n)(x - x^\dagger) \rangle_{\ell^\infty \times \ell^1} + \langle A^* \eta, x - x^\dagger \rangle_{\ell^\infty \times \ell^1} \\ &\leq \mu \|(I - P_n)(x - x^\dagger)\|_{\ell^1} + \gamma_n \|Ax - Ax^\dagger\|_Y. \end{aligned}$$

The triangle inequality yields

$$\|P_n(x - x^\dagger)\|_{\ell^1} \leq \mu(\|(I - P_n)x\|_{\ell^1} + \|(I - P_n)x^\dagger\|_{\ell^1}) + \gamma_n \|Ax - Ax^\dagger\|_Y. \quad (3.2)$$

Now

$$\begin{aligned} &\beta \|x - x^\dagger\|_{\ell^1} - \|x\|_{\ell^1} + \|x^\dagger\|_{\ell^1} \\ &= \beta \|P_n(x - x^\dagger)\|_{\ell^1} + \beta \|(I - P_n)(x - x^\dagger)\|_{\ell^1} - \|P_n x\|_{\ell^1} - \|(I - P_n)x\|_{\ell^1} \\ &\quad + \|P_n x^\dagger\|_{\ell^1} + \|(I - P_n)x^\dagger\|_{\ell^1} \end{aligned}$$

together with

$$\beta \|(I - P_n)(x - x^\dagger)\|_{\ell^1} \leq \beta \|(I - P_n)x\|_{\ell^1} + \beta \|(I - P_n)x^\dagger\|_{\ell^1}$$

and

$$\|P_n x^\dagger\|_{\ell^1} = \|P_n(x - x^\dagger - x)\|_{\ell^1} \leq \|P_n(x - x^\dagger)\|_{\ell^1} + \|P_n x\|_{\ell^1}$$

shows

$$\begin{aligned} &\beta \|x - x^\dagger\|_{\ell^1} - \|x\|_{\ell^1} + \|x^\dagger\|_{\ell^1} \\ &\leq 2\|(I - P_n)x^\dagger\|_{\ell^1} + (1 + \beta)\|P_n(x - x^\dagger)\|_{\ell^1} \\ &\quad - (1 - \beta)(\|(I - P_n)x\|_{\ell^1} + \|(I - P_n)x^\dagger\|_{\ell^1}). \end{aligned}$$

Combining this estimate with the previous estimate (3.2) and taking into account that $\beta = \frac{1-\mu}{1+\mu}$ and $\mu = \frac{1-\beta}{1+\beta}$ we obtain

$$\begin{aligned} \beta \|x - x^\dagger\|_{\ell^1} - \|x\|_{\ell^1} + \|x^\dagger\|_{\ell^1} &\leq 2\|(I - P_n)x^\dagger\|_{\ell^1} + \frac{2}{1+\mu} \gamma_n \|Ax - Ax^\dagger\|_Y \\ &\leq 2\|(I - P_n)x^\dagger\|_{\ell^1} + 2\gamma_n \|Ax - Ax^\dagger\|_Y. \end{aligned}$$

Taking the infimum over all $n \in \mathbb{N}$ completes the proof. \square

Now that we arrived at a variational source condition we summarize some observations the subtle observer can make en route.

Proposition 12. *If A is sequentially weak*-to-weak continuous, the following statements are equivalent:*

- (i) Condition 4 holds for each $\mu \in (0, 1)$,
- (ii) $e^{(k)} \in \overline{\mathcal{R}(A^*)}$ for all $k \in \mathbb{N}$,
- (iii) $\overline{\mathcal{R}(A^*)} = c_0$,
- (iv) A is injective.

Proof. We show (i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (iv) \Rightarrow (i).

(i) \Rightarrow (ii): Fix k , fix $n \geq k$, take a sequence $(\mu_m)_{m \in \mathbb{N}}$ in $(0, 1)$ with $\mu_m \rightarrow 0$ and choose $\xi := e^{(k)}$ in Condition 4. Then for a corresponding sequence $(\eta_m)_{m \in \mathbb{N}}$ from Condition 4 we obtain

$$\|e^{(k)} - A^*\eta_m\|_{\ell^\infty} \leq \|e^{(k)} - P_n A^*\eta_m\|_{\ell^\infty} + \|(I - P_n)A^*\eta_m\|_{\ell^\infty}.$$

The first summand is zero by the choice of ξ and the second summand is bounded by μ_m . Thus, $\|e^{(k)} - A^*\eta_m\|_{\ell^\infty} \rightarrow 0$ if $m \rightarrow \infty$.

(ii) \Rightarrow (iii): $(e^{(k)})_{k \in \mathbb{N}}$ is a Schauder basis in c_0 . Thus, $c_0 \subseteq \overline{\mathcal{R}(A^*)}$. In [7, Lemma 2.1] we find that weak*-to-weak continuity implies $\mathcal{R}(A^*) \subseteq c_0$ and hence also $\overline{\mathcal{R}(A^*)} \subseteq c_0$.

(iii) \Rightarrow (iv): One easily shows that $\overline{\mathcal{R}(A^*)} \subseteq \mathcal{N}(A)^\perp$. Thus, $c_0 \subseteq \mathcal{N}(A)^\perp$. If we have some $x \in \ell^1$ with $Ax = 0$, then for each $u \in c_0 \subseteq \mathcal{N}(A)^\perp$ we obtain

$$\langle x, u \rangle_{\ell^1 \times c_0} = \langle u, x \rangle_{\ell^\infty \times \ell^1} = 0,$$

which is equivalent to $x = 0$.

(iv) \Rightarrow (i): See Corollary 10. □

Note that we have (2.1) if and only if Condition 4 holds with $\mu = 0$. According to Proposition 12 one might start with (2.2) instead of Condition 4. This is an obvious generalization of (2.1). While it is in general not easy to decide whether or not (2.1) holds, (2.2) or, equivalently, injectivity of the operator A can be verified easily. The fulfillment of sequential weak*-to-weak continuity follows in practically all relevant cases from the construction of the problem, see the discussion in the first lines of Section 2.

4 Proof part II: convergence rates

We now collect further proof pieces from the literature to provide the reader with a full proof of Theorem 1. The missing part of the proof is the step from a variational source condition (3.1) to the error estimate in Theorem 1. For a priori chosen regularization parameter we follow the proof of [13, Theorem 1] (while improving constants slightly) and in case of the discrepancy principle we follow the arguments in [6].

In the two proofs we exploit the properties of the function φ in Theorem 1 several times. Simple calculations show that $t \mapsto \frac{\varphi(t)}{t}$ is decreasing. As a consequence we see that $\varphi(ct) \leq c\varphi(t)$ if $c \geq 1$. Both observations will be used without further notice.

Proposition 13. *Let the variational source condition (3.1) be satisfied and choose α in (1.2) such that*

$$c_1 \frac{\delta^p}{\varphi(\delta)} \leq \alpha \leq c_2 \frac{\delta^p}{\varphi(\delta)}$$

with constants $c_1, c_2 > 0$. Then

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^1} \leq \frac{1}{\beta} \left(1 + \frac{1}{c_1} + (1 + 2c_2)^{\frac{1}{p-1}} \right) \varphi(\delta) \quad \text{for all } \delta > 0.$$

Proof. Because x_α^δ is a minimizer of (1.2) we have

$$\begin{aligned} \|x_\alpha^\delta\|_{\ell^1} - \|x^\dagger\|_{\ell^1} &= \frac{1}{\alpha} (T_\alpha^\delta(x_\alpha^\delta) - \alpha \|x^\dagger\|_{\ell^1} - \|Ax_\alpha^\delta - y^\delta\|_Y^p) \\ &\leq \frac{1}{\alpha} (\|Ax^\dagger - y^\delta\|_Y^p - \|Ax_\alpha^\delta - y^\delta\|_Y^p) \\ &\leq \frac{1}{\alpha} (\delta^p - \|Ax_\alpha^\delta - y^\delta\|_Y^p). \end{aligned}$$

and thus the variational source condition (3.1) implies

$$\beta \|x_\alpha^\delta - x^\dagger\|_{\ell^1} \leq \frac{1}{\alpha} (\delta^p - \|Ax_\alpha^\delta - y^\delta\|_Y^p) + \varphi(\|Ax_\alpha^\delta - Ax^\dagger\|_Y). \quad (4.1)$$

Because $\beta \|x_\alpha^\delta - x^\dagger\|_{\ell^1} \geq 0$, we obtain

$$\|Ax_\alpha^\delta - y^\delta\|_Y^p \leq \delta^p + \alpha \varphi(\|Ax_\alpha^\delta - Ax^\dagger\|_Y).$$

If $\|Ax_\alpha^\delta - y^\delta\|_Y \leq \delta$, then the triangle inequality, the properties of φ and the parameter choice imply

$$\|Ax_\alpha^\delta - y^\delta\|_Y^p \leq \delta^p + \alpha \varphi(2\delta) \leq \delta^p + 2\alpha \varphi(\delta) \leq (1 + 2c_2) \delta^p,$$

that is,

$$\|Ax_\alpha^\delta - y^\delta\|_Y \leq (1 + 2c_2)^{\frac{1}{p}} \delta \leq (1 + 2c_2)^{\frac{1}{p-1}} \delta.$$

If, on the other hand, $\|Ax_\alpha^\delta - y^\delta\|_Y > \delta$, then

$$\begin{aligned} \|Ax_\alpha^\delta - y^\delta\|_Y^p &\leq \delta^p + \alpha \varphi(\|Ax_\alpha^\delta - y^\delta\|_Y + \delta) \\ &= \delta^p + \alpha \frac{\varphi(\|Ax_\alpha^\delta - y^\delta\|_Y + \delta)}{\|Ax_\alpha^\delta - y^\delta\|_Y + \delta} (\|Ax_\alpha^\delta - y^\delta\|_Y + \delta) \\ &\leq \delta^p + \alpha \frac{\varphi(\delta)}{\delta} (\|Ax_\alpha^\delta - y^\delta\|_Y + \delta) \\ &\leq \delta^{p-1} \|Ax_\alpha^\delta - y^\delta\|_Y + 2\alpha \frac{\varphi(\delta)}{\delta} \|Ax_\alpha^\delta - y^\delta\|_Y \end{aligned}$$

and thus,

$$\|Ax_\alpha^\delta - y^\delta\|_Y \leq \left(\delta^{p-1} + 2\alpha \frac{\varphi(\delta)}{\delta} \right)^{\frac{1}{p-1}} \leq (1 + 2c_2)^{\frac{1}{p-1}} \delta.$$

In both cases (4.1) can be further estimated to obtain

$$\begin{aligned} \beta \|x_\alpha^\delta - x^\dagger\|_{\ell^1} &\leq \frac{1}{\alpha} (\delta^p - \|Ax_\alpha^\delta - y^\delta\|_Y^p) + \varphi(\|Ax_\alpha^\delta - y^\delta\|_Y + \delta) \\ &\leq \frac{\delta^p}{\alpha} + \varphi \left(\left(1 + (1 + 2c_2)^{\frac{1}{p-1}} \right) \delta \right) \\ &\leq \frac{\delta^p}{\alpha} + \left(1 + (1 + 2c_2)^{\frac{1}{p-1}} \right) \varphi(\delta) \end{aligned}$$

and the lower bound for α leads to

$$\beta \|x_\alpha^\delta - x^\dagger\|_{\ell^1} \leq \frac{\varphi(\delta)}{c_1} + \left(1 + (1 + 2c_2)^{\frac{1}{p-1}} \right) \varphi(\delta). \quad \square$$

Note that in the proof we used arguments similar to the ones in [13], but made changes in the details leading to a better constant in the obtained error estimate. Corresponding estimates in [13, Theorem 1] lead to

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^1} \leq \frac{1}{\beta} \left(1 + 2(2 + p)^{\frac{1}{p-1}} \right) \varphi(\delta),$$

which has a greater constant factor than our estimate. Our estimate with the parameter choice from [13], that is $c_1 = c_2 = 1$, reads

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^1} \leq \frac{1}{\beta} \left(2 + 3^{\frac{1}{p-1}} \right) \varphi(\delta).$$

Proposition 14. *Let the variational source condition (3.1) be satisfied and choose α in (1.2) according to the discrepancy principle (1.3). Then*

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^1} \leq \frac{1+\tau}{\beta} \varphi(\delta) \quad \text{for all } \delta > 0.$$

Proof. Because x_α^δ is a minimizer of (1.2) we have

$$\begin{aligned} \|x_\alpha^\delta\|_{\ell^1} - \|x^\dagger\|_{\ell^1} &= \frac{1}{\alpha} (T_\alpha^\delta(x_\alpha^\delta) - \alpha \|x^\dagger\|_{\ell^1} - \|Ax_\alpha^\delta - y^\delta\|_Y^p) \\ &\leq \frac{1}{\alpha} (\|Ax^\dagger - y^\delta\|_Y^p - \|Ax_\alpha^\delta - y^\delta\|_Y^p) \\ &\leq \frac{1}{\alpha} (\delta^p - \|Ax_\alpha^\delta - y^\delta\|_Y^p). \end{aligned}$$

and taking into account the left-hand inequality in (1.3) we obtain

$$\|x_\alpha^\delta\|_{\ell^1} - \|x^\dagger\|_{\ell^1} \leq 0.$$

The variational source condition (3.1) thus implies

$$\beta \|x_\alpha^\delta - x^\dagger\|_{\ell^1} \leq \varphi(\|Ax_\alpha^\delta - Ax^\dagger\|_Y) \leq \varphi(\|Ax_\alpha^\delta - y^\delta\|_Y + \delta)$$

and the right-hand side in (1.3) yields

$$\beta \|x_\alpha^\delta - x^\dagger\|_{\ell^1} \leq \varphi((1+\tau)\delta) \leq (1+\tau)\varphi(\delta). \quad \square$$

5 Remarks and open questions

In the present paper we only consider the decay of the components of the solution x^\dagger in their natural ordering $x_1^\dagger, x_2^\dagger, \dots$. In [9] a more general formulation was used. There the decay of the components after ordering them by size was considered, which may improve the error estimate for $\|x_\alpha^\delta - x^\dagger\|$ slightly. The same technique can be applied in the present paper, too, but to avoid notational intricateness we did not implement this feature.

Next to the considered a priori parameter choice and to the discrepancy principle also other parameter choice rules lead to the desired error estimate. For example the sequential discrepancy principle or the Lepskii principle can be used (cf. [13]).

An open question is whether the assumption that A is injective can be dropped. The case of non-injective operators is of substantial interest in compressed sensing. Indeed, in [7] convergence rates for ℓ^1 -regularization

were proven without the use of injectivity and also no finite basis injectivity or related properties were assumed. There, source-type conditions quite similar to Condition 4 were required and the question is whether those source-type conditions are always satisfied. Following the ideas of Section 3, Proposition 8 would state

$$\overline{\mathcal{R}(A^*)} = \mathcal{N}(A)^\perp \cap c_0,$$

where $\mathcal{N}(A)$ denotes the null space of A . But the proof of Proposition 9 cannot be carried over directly to the non-injective case. Perhaps, additional assumptions on the ‘angle’ between $\mathcal{N}(A)$ and the faces of the unit ball in ℓ^1 are required.

Another open problem to be solved in future is the interplay between μ and γ_n in Condition 4. We know that there are situations which do not allow $\mu = 0$ (cf. [8]), but on the other hand the condition holds for all $\mu \in (0, 1)$ as we have shown. The γ_n obviously depend on μ and we would like to know more about this dependence. In particular, we do not know whether μ influences the asymptotic behavior of the γ_n if $n \rightarrow \infty$.

Acknowledgments

We thank Bernd Hofmann (TU Chemnitz) for many valuable comments on a draft of this paper and for fruitful discussions on the subject. Research was supported by DFG grants FL 832/1-2, HO 1454/8-2 and HO 1454/10-1.

References

- [1] S. W. Anzengruber, B. Hofmann, and R. Ramlau. On the interplay of basis smoothness and specific range conditions occurring in sparsity regularization. *Inverse Problems*, 29:125002 (21pp), 2013.
- [2] K. Bredies and D. A. Lorenz. Regularization with non-convex separable constraints. *Inverse Problems*, 25(8):085011 (14pp), 2009.
- [3] M. Burger, J. Flemming, and B. Hofmann. Convergence rates in ℓ^1 -regularization if the sparsity assumption fails. *Inverse Problems*, 29:025013 (16pp), 2013.
- [4] I. Daubechies, M. Defrise, and C. De Mol. An Iterative Thresholding Algorithm for Linear Inverse Problems with a Sparsity Constraint.

- Communications on Pure and Applied Mathematics*, 57(11):1413–1457, 2004.
- [5] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Mathematics and Its Applications. Kluwer Academic Publishers, Dordrecht, 1996.
 - [6] J. Flemming. *Generalized Tikhonov Regularization and Modern Convergence Rate Theory in Banach Spaces*. Shaker Verlag, Aachen, 2012.
 - [7] J. Flemming. Convergence rates for ℓ^1 -regularization without injectivity-type assumptions. *Inverse Problems*, 32(9):095001 (19pp), 2016.
 - [8] J. Flemming and M. Hegland. Convergence rates in ℓ^1 -regularization when the basis is not smooth enough. *Appl. Anal.*, 94:464–476, 2015.
 - [9] J. Flemming, B. Hofmann, and I. Veselić. A unified approach to convergence rates for ℓ^1 -regularization and lacking sparsity. *J. Inverse Ill-Posed. Probl.*, 24:139–148, 2016.
 - [10] M. Grasmair. Well-posedness and convergence rates for sparse regularization with sublinear ℓ^q penalty term. *Inverse Probl. Imaging*, 33:383–387, 2009.
 - [11] M. Grasmair. Non-convex sparse regularisation. *Journal of Mathematical Analysis and Applications*, 365(1):19–28, 2010.
 - [12] M. Grasmair, M. Haltmeier, and O. Scherzer. Necessary and sufficient conditions for linear convergence of ℓ^1 -regularization. *Comm. Pure Appl. Math.*, 64:161–182, 2011.
 - [13] B. Hofmann and P. Mathé. Parameter choice in Banach space regularization under variational inequalities. *Inverse Problems*, 28:104006 (17pp), 2012.
 - [14] D. A. Lorenz. Convergence rates and source conditions for Tikhonov regularization with sparsity constraints. *J. Inverse Ill-Posed Probl.*, 16:463–478, 2008.
 - [15] R. E. Megginson. *An Introduction to Banach Space Theory*, volume 183 of *Graduate Texts in Mathematics*. Springer, New York, 1998.

- [16] T. Schuster, B. Kaltenbacher, B. Hofmann, and K. S. Kazimierski. *Regularization Methods in Banach Spaces*, volume 10 of *Radon Ser. Comput. Appl. Math.* Walter de Gruyter, Berlin/Boston, 2012.
- [17] M. Takesaki. *Theory of Operator Algebra I*, volume 124 of *Encyclopaedia of Mathematical Sciences*. Springer, Berlin Heidelberg New York, 2002.